

СТОХАСТИЧЕСКИЙ АНАЛИЗ социальных сетей - методы анализа связей и отношений между акторами (см. Социальная сеть), которые учитывают информацию о распределении структурных характеристик в генеральной совокупности сетей с заданными свойствами (напр., размером) и позволяют проверять гипотезы о форме распределений, значимости параметров модели, пригодности модели для описания данных.

Распределения вероятностей. Как и в общей прикладной статистике, первые методы С.А. были разработаны для параметрического случая, когда известны распределения параметров. Морено (см.) и Х. Дженнингс, а затем У.

Бронфенбреннер в конце 1930-х - начале 1940-х использовали распределение Бернулли для анализа статистических свойств социометрических данных, в частности, для сравнения полученной и предсказанной доли взаимных связей. Появилась возможность проверять статистические гипотезы о том, что связи в сети распределены по закону Бернулли. Начиная с 1950-х для анализа свойств сети и для проверки гипотез привлекают более сложные условные распределения, когда один или более параметров фиксируются. Наибольшее применение получили равномерные распределения при условии, что количество связей в сети равно заданному числу, что количество сделанных акторами выборов равно некоторой постоянной (это особенно полезно для статистического анализа социометрических данных с ограничениями на количество выборов), что количество взаимных связей равно некоторой постоянной и т.д.

Введение условий в распределение позволяет уменьшить число элементов в выборочном пространстве случайных реализаций орграфа заданного размера и тем самым упростить анализ. Для простейших распределений аналитически получены оценки параметров. В более сложных случаях используют метод Монте-Карло или непараметрические критерии.

Диады и триады. Одна из самых продуктивных идей в С.А. состоит в декомпозиции сети на подграфы, особенно на диады и триады, свойства которых хорошо изучены.

Декомпозиция на подграфы позволяет связывать в одной модели микро- и макроструктурные свойства сети и статистически моделировать присущий реальным сетям баланс, дифференциальную популярность акторов, кластеризацию, упорядоченную кластеризацию и транзитивность. Диадой называется неупорядоченная пара акторов i и j с заданными на ней ребрами. Диада принимает четыре состояния: $i \rightarrow j$ и $j \rightarrow i$; $i \rightarrow j$; $j \rightarrow i$; ни $i \rightarrow j$, ни $j \rightarrow i$. Выделяют три изоморфных класса диад: M (от англ.

"mutual", т.е. взаимная; $i \rightarrow j$ и $j \rightarrow i$), A (асимметричная, $i \rightarrow j$ или $j \rightarrow i$) и N (нулевая, связей между i и j нет). Количество диад каждого типа в сети называется переписью диад и используется как многомерная статистика в более сложных методах анализа.

Триадой называется подграф из трех вершин. Каждая триада может принять одно из 64 состояний, принадлежащих к одному из 16 изоморфных классов. Для обозначения типов триад (изоморфных классов) используют обозначения из переписи диад, например, в триаде 003 нет взаимных и асимметричных связей ($M = 0$, $N = 0$) и есть три нулевых связи ($N = 0$). Ниже представлены 6 классов триад, имеющих особенно интересную интерпретацию.

Выбор (choice; $i \rightarrow j$) - это наличие связи между двумя акторами. Этот параметр интерпретируется как показатель плотности сети, насыщенности связей, а также используется для статистического контроля более сложных параметров. Взаимность (mutuality; $i \leftrightarrow j$) - показатель сбалансированности отношений между элементами, тесноты отношений.

Экспансия (2-out-star; $k \leftarrow i \rightarrow j$) - тенденция актора инициировать связи с другими. Популярность (2-in-star; $k \rightarrow i \leftarrow j$) - тенденция других участников сети устанавливать связи с данным актором; экспансия и популярность указывает на доминирование иерархических (однаправленных) отношений в сети.

Посредничество (2-mix-star; $k \rightarrow i \rightarrow j$) - срединное положение элементов, передаточная функция без взаимности; обычно интерпретируется в совокупности с цикличностью. Транзитивность (transitivity; $i \rightarrow j \rightarrow k \leftarrow i$) - одно из самых теоретически наполненных свойств сети, показатель гармоничности связей в зиммелевском смысле. Транзитивными являются отношения, когда друг моего друга становится моим другом. Перепись триад (вектор, состоящий из 16 показателей) рассматривается как многомерная статистика, устойчивая к ошибкам измерения свойств сети, и используется при расчете многих важных характеристик.

Логлинейные (логарифмически-линейные) модели.

Логлинейные модели связывают логарифм зависимой переменной с независимыми переменными. В первой хорошо обоснованной логлинейной модели для социальных сетей, т. н. p_1 (П. Холанд и С. Линард, 1981) в качестве зависимой переменной использовалась диада, принимающая одно из четырех состояний (см. выше), а независимых - эффекты экспансивности и эффекты популярности для каждого актора и два общих параметра - тенденция к образованию связей и тенденция к образованию взаимных связей. Пригодность этой статистической модели оценивалась через функцию

максимального правдоподобия.

Метод p_1 реализован в ряде программ по анализу социальных систем, оценки параметров можно также получить с помощью любой программы логлинейного анализа (после адекватного преобразования данных). Использование модели сдерживалось тем, что она основывалась на нереалистичном предположении о независимости диад, а также тем, что количество оцениваемых параметров росло с увеличением количества акторов (это ограничение снималось в моделях, в которых индивиды разбивались на подгруппы в соответствии с некоторыми характеристиками, например, полом). Одним из существенных и перспективных расширений модели p_1 была попытка использования байесова подхода, учитывающего априорные данные о возможных значениях оцениваемых параметров.

Серьезные новшества случились в методах С.А. в 1990-е. С. Вассерман и Ф. Паттисон предложили т.наз. p^* -модели, позволяющие оценивать сложные статистические модели, в которых связь между акторами рассматривается как зависимая переменная, а структурные свойства социальной сети (характеристики подграфов) - как независимые.

Тем самым большинству описательных методов (см. Анализ социальных сетей) были найдены статистические эквиваленты. p^* -модели основаны на идее марковских случайных графов и их параметрическом представлении, предложенном Франком и Страуссом в 1980-х, а также на методах статистической оценки моделей - псевдомаксимальном правдоподобии.

Используя теорему Хаммерсли-Клиффорда о том, что случайный оргграф имеет вероятность, зависящую только от полных подграфов графа зависимости, Страусс и Икеда предложили лог-линейную модель для оценки этой вероятности, а Вассерман и Паттисон - логит-модель. Зависимой переменной служит логит, или логарифм отношения шансов (условных вероятностей того, что актер i выбирает актора j , и того, что i не выбирает j), а независимыми - структурные свойства графа. В качестве таких структурных параметров используются статистики для триад, поскольку их стохастические свойства известны, и они имеют интересную содержательную интерпретацию. Например, замкнутые транзитивные триады ограничивают возможности индивидуального действия и снижают вероятность конфликтов. p^* -модели могут оцениваться с помощью любой программы логистической регрессии, но перед этим социоматрицу следует преобразовать, т.е. рассчитать триадические и т.п. показатели.

Предварительные преобразования выполняют программа `prepstar` или командные

файлы для SPSS, доступные в Интернете. Другие программы - PSPAR, STOCNET - позволяют оценивать модели непосредственно. r^* -модели оценивают неслучайность связей в социальных сетях (баланс, кластеризацию, упорядоченную кластеризацию и транзитивность).

Стохастическое блочное моделирование. В литературе представлены два основных подхода. В первом задается детерминистская (нестохастическая) блочная модель (см.

Анализ социальных сетей), на ее основании предсказываются связи, а затем реально полученные и предсказанные связи сравниваются по одному из критериев пригодности модели. Обычно для этой цели используют непараметрические критерии, напр., перестановок, или аналог квадрата множественной корреляции, известный из регрессионного анализа. Этот подход не является статистическим в прямом смысле слова. В рамках второго подхода сначала делаются предположения о взаимодействии в диадах, затем постулируется стохастическая блочная модель. Критерием стохастической эквивалентности служит неизменность распределения вероятностей при замене одного эквивалентного параметра другим.

Для оценки моделей используется критерий, основанный на функции максимального правдоподобия. Примером развития идеи стохастических блоков служит недавняя модель К. Новицкого и Т. Снайдерса, использующая модель латентных классов. Актуальной методологической задачей является объединение стохастической блочной и r^* -моделей.

С.В. Сивуха